By the Numbers

The Newsletter of the Statistical Analysis Committee of the Society for American Baseball Research
Volume 4, Number 2 June, 1992

Committee News

Convention Issue. This is the convention issue of ***By the Numbers***; I hope all of you who are on the mailing list got it *before* you came to St. Louis. I hope some of you who are new readers will be interested enough to join us as we seek to examine baseball issues by using statistical analysis.

Statistical analysis--the use of more or less formal statistical techniques to examine questions--is not the only way to learn about baseball, but it is one way to increase our knowledge. Understanding what our authors have to say will require some effort (more in some cases than in others), but I think you will find the rewards worth the effort.

This issue contains an article by John Strkyer looking at the importance of achieving *more* of something (hits, doubles, men on base, stolen bases, etc.) in a game. He undertook this to extend a report on the "Numbers" page in ***Inside Sports*** that suggested that teams stealing more bases in a game than their opponents were very likely to win. Read John's article to find out how a team is *most* likely to win.

Bruce Cowgill has written a piece on relative performance measurement. An understanding of this technique will make comparisons of players from different eras easier, if still not completely smooth. Vince Coleman may have been 12th on the all-time stolen base list at the end of 1991, but he looks much better after adjusting for relative performance levels.

David Smith looks at the "big bang" theory of offense--a winning team is likely to score more runs in one inning that the losing team scores in the game. Maybe the "big bang" isn't as useful a theory as we might think.

John Benson takes a look at "pitching inside," using the ratio of hit batters to unintentional bases on balls. What does a *high* ratio mean? And why does this tell us that Steve Howe may have good control even though he hits a lot of batters (I'll avoid the temptation to take cheap shots about hits here).

Finally, I have wrtten a piece looking at whether Hall of Fame selections are predictable, and what sorts of performances seem to matter. Want to know the "best" and "worst" HOF selections of all time? Read on.

Inside Sports. Larry Burke of ***Inside Sports*** magazine called to say they plan to expand their "Numbers" page (those of you who've been with us for a while know I've taken some shots at what has appeared there). He asked me to publicize their interest in acquiring more--and more interesting--statistical nuggets for that page. If you are interested, send your statistical nuggets to

Larry Burke
Inside Sports
990 Grove Street
Evanston, IL 60201-4370

Milestones and Memories. I have now received two or three copies of ***Milestones and Memories***, edited by Jim Fredlund. It's not at all analytical, but it is filled with interesting records and trivia. If you are interested in receiving it, write Jim for subscription information at

Jim Fredlund
Milestones and Memories
P. O. Boc 679
Jessup, MD 20794

Extra Innings. Bill French has produced an extensive comparison and analysis of varying measures of offense, in a newsletter called ***Extra Innings***. I think it's an excellent piece of work and urge you to write Bill to acquire it (send him money, however--it's a 10-page effort; $3.00 will cover his costs):

Bill French
1221 Stanford
Oakland, CA 94608

Convention Meeting. Those of you who get this before the commmittee's meeting in St. Louis might consider dropping by on Sunday morning at 9 o'clock. We'll talk about what the committee does, what you can do, and where the future will lead us. One item that we will certainly talk about is a transition in the committee chairmanship. Following this issue, Rob Wood and I will be co-chairing the committee for a while, after which we expect he will take over entirely. He will take over the newsletter immediately, so future newsletter submisions should be directed to Rob. His address and mine are both shown below.

The Next Issue. The next issue is the convention follow-up issue. We plan to provide a forum for some of the research presented in St. Louis, so look for that in September. Beyond that, get your material to Rob.

Donald A. Coffin
Indiana University Northwest
3400 Broadway
Gary, IN 46408

Rob Wood
2101 California St., #224
Mountain View, CA 94040

****

Box Score Category Domination

By John Stryker

In the last issue of ***By The Numbers*** it was pointed out that a recent magazine article said that teams that stole more bases than their opponents had a great winning percentage. Because there are all sorts of things that cloud that issue, I decided to look at nearly every boxscore category and see just how significant each was in terms of winning percentage.

A database of all 1991 National League games was used. I checked twenty-one box score categories. For each category, I looked at every game with two questions: 1) Did one team dominate that category, i.e. the category wasn't tied; and 2) If so, did that team win or lose. I created five composite categories also relating to the stolen base question, checking opportunutnies and success rate.

The resulting table is shown in the next column for your inspection and interpretation. My comments follow the table.

	W	L	WPCT
AB	479	395	0.548
R	970	0	1.000
ER	867	33	0.963
H	683	183	0.789
BB	544	282	0.659
K	379	484	0.439
RBI	903	14	0.985
E	210	408	0.340
LOB	505	362	0.582
GDP	288	340	0.459
2B	463	254	0.646
3B	208	110	0.654
HR	422	172	0.710
SB	399	220	0.645
CS	259	202	0.562
SH	337	159	0.679
SF	258	121	0.681
HBP	172	115	0.599
WP	154	243	0.388
BK	57	70	0.449
PB	52	82	0.388
BR[1]	732	166	0.815
M1B[2]	663	224	0.747
SB%[3]	409	264	0.608
SBA%[4]	460	381	0.547
SBS%[5]	410	314	0.566

There were 970 (of a scheduled 972) games played in the National League in 1991.
1. BR = H+W+HP
2. M1B = BR-2B-3B-HR
3. SB% = SB/(SB+CS)
4. SBA% = (SB+CS)/M1B
5. SBS% = SB/M1B

1) My calculations for stolen bases (the magazine category) differ slightly from what was published in the last ***BTN***. There may still be some minor errors in my database; I don't think they would alter any conclusions we may draw from these numbers.

2) The importance on OB% is again emphasized. The equivalent category here (H+BB+HBP) wins more than four out of five games.

3) Most things we look at in the boxscore are positive. Of the twenty-one categories checked, only six (K, E, GDP, WP, BK, PB) led to a losing record among dominant teams. A seventh "negative" category, LOB, had a fine winning percentage. Apparently teams

that leave more men on often have baserunners to spare.

4) On the stolen base issue, the numbers essentially check out. But note that teams that get caught stealing more than the opposition are pennant contenders as well. Perhaps the players likely to attempt steals, being more athletic in general, use their speed in other areas (read: defense, taking extra bases, etc.) to help their team win. Note also that teams who do well stealing do not win as much as those who just have more men on first base (appoximately equal to stolen base opportunities). In any case, if the point was that one should build a team with base-stealers, these numbers refute that. Home runs prove to be a stronger indicator, and merely reaching base far stronger than that. But we knew that.

5) I am intrigued by the similarity of SH and SF. Each is basically the trading of a run for an out. This may relate to the value of a single run to winning the game (any ideas out there, folks?). Naturally, the winning percentage of teams that score the most runs is 1.000, and darn near it for those with the most RBI and ER. [Note - ER is as calulated by individual opposing pitcher, not by opposing team (see rule book for the distinction).]

I am interested in any comments you may have. Please address them to John Stryker, PO Box 1433, Northbrook, IL 60065-1433.

****

Relative Performance Measurement II: A New Technique

By Bruce W. Cowgill

In ***The Hidden Game of Football*** (Carroll, Palmer, & Thorn: 1988) a major issue is raised for football, an issue that has been much researched in baseball. The issue is how to compare players across different eras. The team of Palmer and Thorn tackled the baseball question in ***The Hidden Game of Baseball*** (1985). Interestingly, the analysis they used to compare football players was not used for baseball players. The obvious reason is that baseball statistics can be studied on an individual basis, whereas football is a much more team-oriented sport. Therefore, this analysis was not needed for baseball.

Before I explain this new measurement, a similar measurement appeared in ***The Baseball Research Journal*** (1990). Ron Skrabacz's paper "Relative Performance Measurement" evaluates players on the basis of a players appearance on the list of league leaders. He states: "RPM is nothing more than the measurement of a player's performance relative to his peers within his league. It ultimately measures how dominant a player was or is during his era, and then allows comparisons with other players, regardless of era." This form of measurement takes into account scoring rules, equipment, ballparks, and players of each era. Simply put, the goal is to find out "how much" the greatest of one era exceeded the others of his time. This is the basis of the new measurement.

I have attempted to apply the analysis used in The Hidden Game of Football to baseball. This analysis (I will call it RPM2) like Skrabacz's RPM is based on league leaders. However, unlike RPM, the only position on the list that matters is the leader himself. The leader in any category is given 100 points, and all subsequent players are rated as a percentage of the leaders total. The percentage is multiplied by 100 for simplicity. This is done for the players entire career. Then, a weighted average is taken each year to get a career average that can be compared to other players (see Boggs example). Note that a perfect RPM2 rating for a career is 100 and is only obtainable if the player leads every year.

Contrary to Skrabacz's RPM, the distance from the leader does matter. He cites Fred McGriff's 1989 home-run title over Joe Carter by just one homer as being no different than Ruth's 1920 title over Sisler by 35. He rates these two incidents the same -- I disagree. Ruth accomplishment was much greater than McGriff's (just ask George Sisler). This measurement would give Ruth 100.0 points and Sisler 35.2 points compared to McGriff's 100.0 points and Carter's 97.2 points. By definition, we are not giving any bonus to the leader who wins the title by a large amount, but we are compensating the subsequent finishers who stay in the race to the end. So, a Ricky Henderson who loses the batting title to an I'm-going-to-sit-out-and-collect-my-award George Brett is not penalized.

Batting average is the most notable statistic that shows a dominance of past players over modern-day players. Thorn and Palmer's Relative Batting Average (RBA) tried to show this by comparing league batting averages to players batting averages. Looking at the current list of leaders, only Boggs from the past 15 years shows up in the top-25 lifetime batting averages; however, five players from the past 15 years show up in the top-25 lifetime RBA. Only Brett shows up in the top-50 season averages, but five players show up in the season RBA.

Table 1: Wade Boggs RPM2

YEAR	AB	HITS	AVG	RATE
1982	338	118	.349	100.0
1983	582	210	.361	100.0
1984	625	203	.325	94.8
1985	623	240	.368	100.0
1986	580	207	.357	100.0
1987	551	200	.363	100.0
1988	584	214	.366	100.0
1989	621	205	.330	97.3
1990	619	187	.302	91.8
1991	546	181	.332	97.4
TOTALS	5699	1965	.345	97.99

I disagree that past ballplayers are more dominant than today's with a few exceptions. For example, few would agree (including myself) that Boggs is the best hitter of all-time. I am not talking power or any other performance measure just hitting. Most people pick Cobb or Williams. However, this analysis reveals that Boggs is ahead of both of them. That is, Boggs dominated his time like no other hitter has before him. Granted, he has had a short career so far, but to date this has been true. Furthermore, some will argue that today's ballplayers are not as good of hitters as in the past. This does not always sit well with old-timers, but the average athlete today is better than the average athlete of yesterday. Case in point, note the track and field records that have been broken in the past ten years. This may be stretching, but if the athletes are better, then we can assume that ballplayers are better (on average). If this assumption is true, then this type of measurement has greater explanatory power.

This analysis does have some shortcomings. RPM2 should only be used as a career measurement not as a seasonal comparison. Furthermore, I, due to time constraints, analyzed only the current lifetime leaders. Thus, our rankings of RPM2 are just based on the top-5 or 10 in a category. It is likely that players who did not rank in the top-5 or 10 on the base statistic would rank on the RPM2 leader board. An example of this is with Vince Coleman who currently is 12th in career stolen bases but rates a very high 96.76 RPM2. Ideally, a Pete Palmer database could give leaders in RPM2 for each relevant statistic. The largest problem I had was in calculating RPM for batting averages. Because of the rules for winning a batting title (currently 3.1 plate appearances per team game), several players had batting averages that were high enough to win titles but did not have the necessary number of at bats or plate appearances. For these cases, I ran two separate tests. In the first test, I awarded 100 points as if they would have lead. For the second test, I awarded the player 0 points. Actually, the figure should somehow compensate them for doing well, but not as much as the leader. I denote the second case as RPM2/A and included the results in the table. I will leave it up to you to judge where the hitter should be placed within this max/min interval.

One final problem is the home park effect. I became suspicious when Boggs and Williams turned up number one and two, respectively. I am not sure what to do about this, so keep in mind where these players performed their respective feats. I did include Batting Runs/Adjusted for Park Effects for some comparison.

Note: Rounding errors may occur.
Statistics taken from Total Baseball (1990).
Leaders determined by rules specific to that time period.

Table 2: Lifetime RPM Batting Averages

A: Batting Average

Rank	Player	BA	Titles
1	Cobb	0.366	10
2	Hornsby	0.358	7
3	Jackson	0.356	0
4	Delahanty	0.346	2
5	Boggs	0.345	5
6	Speaker	0.345	1
7	Williams	0.344	7
8	Haminton	0.344	2
9	Brouthers	0.342	5
10	Ruth	0.342	1

B: RPM2

Rank	Player	RPM2
1	Boggs	97.99
2	Williams	96.46
3	Hornsby	96.00
4	Cobb	95.63
5	Brouthers	92.02
6	Jackson	91.26
7	Hamilton	90.18
8	Delahanty	89.63
9	Ruth	89.01
10	Speaker	88.80

C: RPM2/A

Rank	Player	RPM2/A
1	Hornsby	95.94
2	Cobb	92.61
3	Boggs	92.06
4	Brouthers	92.02
5	Williams	90.99
6	Hamilton	90.18
7	Jackson	89.75
8	Delahanty	89.63
9	Ruth	89.01
10	Speaker	88.80

Table 3: Home Runs

A: Home Runs

Rank	Player	HRs	Titles
1	Aaron	755	4
2	Ruth	714	12
3	Mays	660	4
4	Robinson	586	1
5	Killebrew	573	6

B: RPM2

Rank	Player	RPM2
1	Ruth	92.19
2	Killebrew	84.74
3	Aaron	82.40
4	Mays	78.93
5	Robinson	72.12

Table 4: Runs Batted In

A: Runs Batted In

Rank	Player	RBI	Titles
1	Aaron	2297	4
2	Ruth	2209	6
3	Gehrig	1990	5
4	Musial	1951	2
5	Cobb	1937	4

B: RPM2

Rank	Player	RPM2
1	Gehrig	87.72
2	Ruth	86.53
3	Aaron	82.00
4	Musial	76.00
5	Cobb	73.17

Table 5: Batting Runs			
A: Batting Runs			
Rank	Player	BR	Titles
1	Ruth	1322	9
2	Williams	1166	10
3	Cobb	1032	7
4	Musial	983	8
5	Gehrig	918	4
B: RPM2			
Rank	Player	RPM2	
1	Williams	94.02	
2	Ruth	91.76	
3	Musial	87.67	
4	Gehrig	86.50	
5	Cobb	80.98	

Table 6: Adjusted Batting Runs			
A: Adjusted Batting Runs			
Rank	Player	ABR	Titles
1	Ruth	1355	10
2	Williams	1093	8
3	Cobb	1018	8
4	Gehrig	966	4
5	Musial	930	9
B: RPM2			
Rank	Player	RPM2	
1	Ruth	92.21	
2	Williams	91.66	
3	Musial	88.50	
4	Gehrig	88.45	
5	Cobb	81.46	

Table Stolen Bases			
A: Stolen Bases			
Rank	Player	SB	Titles
1	Henderson	994	11
2	Brock	938	8
3	Hamilton	912	5
4	Cobb	891	6
5	Collins	744	4
12	Coleman	586	6
B: RPM2			
Rank	Player	RPM2	
1	Henderson	96.99	
2	Coleman	96.76	
3	Hamilton	87.41	
4	Brock	86.72	
5	Cobb	81.47	
6	Collins	74.38	

****

The Big Bang: A Big Bust?

By David W. Smith

It has become fashionable in baseball circles in recent years to talk about the significance of the "big bang." Popularized by sportswriter Tom Boswell--with a big assist from Earl Weaver, who built his orioles around the three-run homer--the big bang theory is now used to explain both the origin of the universe as well as the decisive moment in most baseball games.

By definition, a big bang occurs in a baseball game whenever the winning team scores more runs in a single inning than the losing team does in the entire game. (We'll leave cosmology to the physicists and astromoners.) Thus, all shutouts are big bangs and could probably be eliminated from the calculation. League summaries for the 1991 season and for the entire 1984-1991 period are shown in Table 1.

This study was prompted by a desire to test the assertion by some sportscasters (e.g., Harry Kalas of the Phillies) that the big bang is a common event and therefore important in understanding the game. On a superficial level, such an assertion is correct--nearly 50%

of all games are indeed big bangs. However, when we try to decide what the significance of that 50% figure is, it becomes evident that the chance of winning with a big bang is much more directly related to a team's ***defense*** (its ability to hold the opponents to fewer than three runs) than it is to the team's offense.

Table 1: Big Bangs and Shutouts		
	1991	1984-1991
American League		
Games	1134	9062
Big Bangs	530	4186
Percent	46.7%	46.2%
Excluding Shutouts		
Games	984	8015
Big Bangs	380	3139
Percent	38.6%	39.2%
National League		
Games	970	7760
Big Bangs	453	3617
Percent	46.7%	46.6%
Excluding Shutouts		
Games	848	6734
Big Bangs	331	2591
Percent	39.0%	38.5%

There is an unfortunate tendency in baseball analysis to see a correlation such as this and jump to a conclusion about cause and effect. However, direct evidence to support a cherished hypothesis is usually harder to come by. In the present case, we can express the dilemma in the form of a question: "Does scoring many runs in a game lead to a greater tendency to have big innings, or does the greater likelihood of a big inning automatically mean that the team wills core more runs in a game?"

In it instersting that the percentages of big bangs in the two leagues are so similar (see Table 1), given the presence of the DH in the American League (and thus higher scoring in the AL). These similarities lead us to consider what the big bang is supposed to represent. Sine it is usually seen as a sign of the "big inning," the numbers should be examined to see if there is any merit in what we might call the "Earl Weaver Method."

To that end, it is interesting to note that the very large majority of big bangs occur when the losing team scored 2 runs or fewer, as shown in Table 2.

Table 2: Percentage of Big Bangs in Which Losers Score 0, 1, or 2 Runs

			American League		
Year	BB	0	1	2	%
1991	530	150	152	119	79.4%
1984-1991	4186	1047	1345	950	79.8%

			National League		
Year	BB	0	1	2	%
1991	479	122	165	101	81.0%
1984-1991	3617	1026	1179	812	83.4%

Again we see similar percentages over the entire 8-year period, with the NL having a slightly higher frequence of occurrence. This overwhelming occurrence of big bangs in games where the losers score fewer than three runs leads to the conclusion that the big bang is not really a measure of a big offense, but an incidental consequence of a well-pitched game. Add on the general perception that the NL is the "pitcher's league" and the conclusion is even stronger.

As I noted at the beginning, all shutouts are big bangs. Table 3 gives the chance of a big bang when a team ***allows*** 1 run, 2 runs, or more.

Some interesting differences are emerging between the two leagues as well, as the percentages in the last two categories are lower in the NL. This difference presumably reflects the overall lower scoring in the NL--in 1991, the AL averaged 8.9 runs per game (both teams), compared to 8.2 runs per game in the NL.

Since it is clear that most big bangs occur when the losers score fewer than 3 runs, it is useful to consider the general chance of winning in all games where a team allows fewer than three runs, whether it is a big bang or not. Table 4 gives the appropriate numbers.

Table 3: Chance of a Big Bang as a Function of Runs Allowed				
American League				
Year	Runs Allowed	Games	Big Bangs	%BB
1991	1	202	152	75.2%
	2	229	119	52.0%
	3+	553	109	19.7%
1984-1991	1	1735	1345	77.5%
	2	1794	950	53.0%
	3+	4486	844	18.8%
National League				
Year	Runs Allowed	Games	Big Bangs	%BB
1991	1	217	165	76.0%
	2	214	101	47.2%
	3+	417	65	15.6%
1984-1991	1	1583	1179	76.0%
	2	1709	812	47.5%
	3+	3442	600	17.4%

Table 4: Won-Loss Records When a Team Allows Fewer than 3 Runs			
AL			
Year	Wins	Losses	Pct.
1991	581	98	0.856
1984-1991	4576	701	0.867
NL			
Year	Wins	Losses	Pct.
1991	553	102	0.844
1984-1991	4318	783	0.847

Again we have the conclusion that the big bang isn't really measuring a big offensive performance. Big bangs predoinantly occur when the winner doesn't allow many runs.

Where then does the appeal of the big bang theory come from? Certainly a big outburst of scoring in an inning is dramatic, and a large lead may be good for a manager's digestion. Nevertheless, scoring, say, 6 runs in an inning during a game in which the opponents are held to 2 runs is hardly a meaningful indication os the value of an overpowering offense. How often does a "big inning" actually occur?

If we define a big inning as one in which a team scores 4 or more runs, then it is a surprisingly rare event, as shown in Table 5 (above). Again, the NL shows slightly lower scoring than the AL, but the patterns are similar in both leagues and demonstrate that the "big inning" is not a common event. Returning to the definition I offered of a big inning (scoring 4+ runs), then it seems unreasonable to base your strategy on an event that occurs less than 2% of the time.

We certainly know that different teams have different strategies, depending on their personnel and home parks (hit-run, stolen bases i n Busch Stadium; bombs away n Fenway Park). Therefore, it is reaosnable to expect a significant variation between teams in the ability to put a big inning on the board. As shown in Table 6, this expectation is met, with Texas, Milwaukee, and Detroit leading the way, while Montreal trails badly.

Table 5: Number and Percentage of Innings With Different Numbers of Runs			
AL Year	Runs/ Inning	Number of Innings	Pct.
1991	0	14882	72.8%
	1	2994	14.6%
	2	1369	6.7%
	3+	499	2.4%
1984-1991	0	117625	72.5%
	1	24610	15.2%
	2	11293	7.0%
	3+	3985	2.5%
NL Year	Runs/ Inning	Number of Innings	Pct.
1991	0	12857	73.7%
	1	2592	14.9%
	2	1140	6.5%
	3+	344	1.9%
1984-1991	0	103257	73.8%
	1	20885	14.9%
	2	9073	6.5%
	3+	2849	2.0%

As with our other scoring measures, we find the top of this list is dominated by AL teams and the bottom has mostly NL teams.

Table 7 is a list of the number of big bang wins and losses by team in the last 8 years.

Table 6: Percentage of innings in Which a Team Scores 4+ Runs (1991)

Team	%	Team	%
Texas	3.3%	Seattle	1.9%
Milwaukee	3.1%	Philadelphia	1.9%
Detroit	3.0%	Los Angeles	1.9%
White Sox	2.9%	Batlimore	1.9%
Kansas City	2.8%	Houston	1.8%
Pittsburgh	2.8%	Cincinnati	1.8%
Oakland	2.6%	San Francisco	1.7%
Minnesota	2.5%	Yankees	1.7%
California	2.4%	Cleveland	1.7%
Mets	2.2%	Toronto	1.6%
Boston	2.2%	Cubs	1.5%
Atlanta	2.2%	Expos	1.0%
San Diego	2.0%		
St. Louis	2.0%		

Table 7: Big Bang Wins and Losses, By Team, 1984-1991

Team	Wins	Losses	Team	Wins	Losses
NYN	359	240	DET	299	280
TOR	357	234	NYA	298	292
LA	348	308	CHIA	291	311
KC	332	310	SF	290	300
SD	329	292	PIT	289	299
BOS	324	269	MIL	286	309
HOU	318	332	BAL	275	323
MIN	316	318	SEA	271	329
OAK	315	268	CHIN	269	303
STL	313	286	PHI	260	319
CAL	310	299	TEX	259	324
CIN	305	309	CLE	253	321
MON	300	287	ATL	237	342

Most Wins in One Season:
Los Angeles, 1985 57
Most Losses in One Season:
Baltimore, 1988 60

In closing, we should ask: Is there any value to the notion of the big bang? It led us to some interesting conclusions about scoring, but it doesn't seem to offer any clearcut indication of which team is going to win. As many analysts have noted over the years, a winning team requires a balance between scoring runs and preventing its opponents from scoring. Simply addressing the offensive half or the defensive half of the equation in isolation cannot be expected to give us the whole picture.

****

Steve Howe's Favorite Statistic: The Pitcher's "Inside Ratio"

By John Benson

Rain is a friend of baseball journalists; it often creates no-hurry interview situations. during a March downpour, I stepped into the groundskeeper's shed at Jack Russell Stadium, and found Steve Howe. We already knew each other, so it wasn't like running into a stranger. (Howe was the only player in 1991 who asked me what his Rotisserie price should be in 1992. "Get two saves in Florida," I told him, "and you'll be a $15 pitcher; outherwise about $5." He sold for $15 in my AL this year.)

On this occasion we chatted about statistics in general and ratios in particular. Howe's favorite pitcher ratio is hit batsmen divided by unintentional walks [HB/(TBB-IBB)]. The short name for this is Inside Ratio (IR). Howe led the major leagues in IR by a huge margin last year (and in 1987, the last time he appeared in the majors, if you want to look it up). "You have to pitch inside," he says. "You got to. Sometimes you hit a batter because you're wild. But if you look at walks, you can see if a pitcher is wild or not. Last year I hit three batters, and walked five [unintentionally]. That's a great ratio. I wasn't throwing at anyone. I was just pitching inside. I need the outside corner to get people out. That's the toughest pitch there is." He pointed a bat at a sot knee high and as far away as the bat would reach. "That spot is just about impossible for a hitter and it's a strike."

Before launching into the subject of IR, I want to assure all readers that I understand the statistical insignificance of numbers like 3 and

5, taken out of context. In fact, my first reaction to this stat, after Howe piqued my interest, was t look at some bigger numbers, team totals in particular. I expected to find that, over the course of 162 games and more than 6000 batters faced, the IR for every team would be about the same. But I was also aware that coaches and managers can affect pitching style, so if there was anything such as IR proclivity, team totals would show it.

Guess what? Oakland led the league in IR in 1991 with an 0.088 ratio. Their pitching coach is Dave Duncan, whose motto is "Pitch inside." Duncan is frequently mentioned as a potential major league manager, because he knows about a hundred other useful mottoes, and doe shis homework, too. The worst IR in the AL last lear belonged to the Tigers, who porduced a 0.047 ratio, barely half of Oakland's. I must confess I had to look it up: Who is the pitching coach in detroit, anyway? Ah, Billy Muffett. If he's being considered for management anywhere, I am not aware of it. In the National League in 1991, St. Louis had the highest IR with 0.116, while San Diego was way down at the bottom at 0.032. So I concluded IR can tell you SOMETING.

Applied to any individual player in any one season, the numbers are tiny, and the discussion quickly becomes anecdotal. You could look at Rob Dibble's IR in 1991 (0.000) and say, "Oh, that's why he had trouble." Tue, Dibble hit one base*runner* and one fan last year, but no batters. And he admitted that he was afraid to pitch inside, mainly because of official warnings. You could also look at baltimore superprospect Arthur Rhodes in 1991 (0.000) and say, "Oh, he's got to pitch inside more to be successful."

At the high end of the spectrum, you find some successful pitchers with an IR above 0.200. Take a look at what happend when we average the IR for every pitcher on my list of above-average Rotisserie selections for 1992 and compare it with the average IR for my list of "bad" pitchers--those who will hurt their Rotisserie rosters if they pitch like I expect them to (stats are for 1991). These are not the worst pitchers in baseball; they are good enough to get a substantial number of innings this year. They are classified as bad, in this essay, because they offer the probability of a poor ERA and a high OBP for opponents, combined with enough IP to damage their teams in this category.

The overall IR for the "bad" group is 0.064, compared to 0.076 for the "good" group. This difference is not quite statistically significant, although it may have "baseball" significance; the difference is about as large as the difference between the two leagues: 0.062 (NL) and 0.075 (AL) for 1991.

Obviously, working inside is just one tool available to a pitcher. And it isn't such a powerful tool that it can make up for weakness in other areas. It is doubtless better to have a 95 MPH fastball abd good controll, than to have a world of knowledge about pitching inside. Just the same, the next time I have to choose between two pitchers for my Rotisserie roster, I'm going to look at their IR's, and remember Steve Howe's remarkable 1991 season.

(John Benson is the editor and publisher of ***John Benson's WINNING ROTISSERIE BASEBALL Monthly***, from which this essay is reproduced with his permission. If you are interested in subscribing to the ***Monthly***, write Diamond Analytics Corporation, Wilton Center, P. O. Box 7302, Wilton, CT 06897; subscriptions are $59 for one year, $99 for two years, or $35 for six months.)

****

A Model of the Hall of Fame Selection Process

By Donald A. Coffin

We are unlikely ever to agree completely on who *ought* to be in the Baseball Hall of Fame. We can, however, look at the selection process to determine whether there are regularities in it--whether the selection process is consistent and predictable. If it is, we can then begin to understand the implicit standards for selection to the Hall of Fame, and to understand why some players, who might seem otherwise highly qualified, remain outside the Hall. We can also look more critically at the qualifications of candidates for the Hall of Fame, and we can perhaps make some suggestions about qualified candidates who have been overlooked.

One of the difficulties in doing all this is that players have been selected to the Hall of Fame by two different bodies. The more widely understood process is the vote of the Baseball Writers of America, in which a 75% favorable vote is required. The second process is reconsideration by the Veterans; Committee.

A second difficulty is that the rules change. The principal rules change recently is the new requirement that the Veterans' Committee can consider only players who have received at least a 60% favorable vote at some time from the BBWA. Another important rules change is that a player must receive at least 5% of the votes cast in order to be retained on the ballot.

A third difficulty is that, over time, the relevant standards for Hall of Fame membership change. This is most dramatically evident in the consideration of pitchers, with the increasing importance of relief pitchers. I think it is likely that we do not yet fully understand how to evaluate relief pitchers for Hall of Fame membership, and that models of selection of pitchers are likely to be difficult to develop until we do understand more clearly how to evaluate relief pitchers. Changing standards also affects other players as well. For example, in the 1930s, the average batting average of retiring players with extended careers (1500+ games) was over 0.300. So an player, for example, who had a batting average of 0.300 at retirement was, in fact, a below-average hitter for average. Unless we can track the effects of changes in performance standards for hitters as well as for pitchers, we are unlikely to model the selection process very well, either.

The purpose of my analysis has been to examine the implicit selection standards for non-pitchers, to see whether the standards for selection of players to the Hall of Fame have been consistent. I do not intend to consider whether those standards are "correct," or "rational," or the standards I would use. If the standards have--or have not--been consistently applied, then we will add to our knowledge of Hall of Fame membership.

Measuring Performance. Our first task is to develop measures of performance that can be used to model the Hall of Fame selection process. I have developed two offensive measures and I have also used one defensive measure. The offensive measures provide an indication of by how much a player exceeds his contemporaries in offensive performance. To do this, I first identified the group of players I though was relevant and then defined a set of retirement cohorts.

The relevant set of players is composed of all players who appeared in at least 1500 games during their careers (1200 for catchers) or who have been selected to the Hall of Fame. These players are clearly among the best ever to play major league baseball--their extended careers are evidence of that. We are, therefore, asking who among the best players ever has been selected to the Hall of Fame.

I then identified retirement cohorts, in order to compare players to their contemporaries, rather than trying to compare players from eras in which performance levels are substantially different. I used five-year periods to define these retirement cohorts, beginning with the 1925-1930 period (my only six-year period), and then proceeding 1931-1936 up to 1986-1990. In general, the number of players in a retirement cohort has tended to rise (after a drop in the 1951-1955 and 1956-1960 periods).

The first offensive measure is what I call the Type I measure. It measures a player's career totals in eight offensive categories, compared to his contemporaries. The eight categories are hits, doubles, triples, home runs, runs scored, runs batted in, walks, and stolen bases. To calculate the Type I measure, I first calculated the means (x_j) and standard deviations (σ_j) for each category. I then calculated, for each category, a Z-score for each player, as follows:

$$(1) \qquad Z_{ij} = (X_{ij} - x_j)/\sigma_j$$

where X_{ij} is player i's performance in category j. This tells us how many standard deviations above or below the mean each player is in each category. For example, if the average number of hits made by members of a retirement cohort is 2000, if the standard deviation of hits is 200, and if a player achieves 2600 hits, then he is three standard deviations above the average in hits ($Z_{ij} = 3.0$).

After calculating each player's Z_{ij} for all categories, I then summed the Z_{ij}'s for all eight categories. This is the Type I measure for each player. It measures how much above (below) average he is in all offensive categories combined. The average Z_{ij} is 0 in each category, so the average Type I score is also 0.

Consistently, in each of the retirement cohorts, the standard deviation of the Type I measure was about 6. I therefore identified players with outstanding career *totals* as being those players for whom the Type I score was 6.0 or greater; about 1/6 of all players in my sample would, in general, fall in this category.

My Type II measure is similar, except it focuses on career *averages*. Here, I used three career averages--batting average, extra base power (SA-BA), and walks per plate appearance [(Walks)/(Walks+At-Bats)]. Had I used slugging average and on-base average as the second and third categories, I would, in effect, have been counting batting average in all three measures. I calculated Z-scores for these categories, as I did for Type I measures, and then summed the Z-scores to get a Type II score. Again, the average Type II score was zero. In each retirement cohort, the standard deviation of the Type II score was about 2, so I identified players with outstanding career *averages* as being those players for whom the Type II score was 2.0 or greater. Again, about 1/6 of the players in my sample would be identified as having outstanding career averages.

Having calculated both Type I scores and Type II scores for all players in my sample, I was then able to classify players into one of four offensive groups. **Type A** players have outstanding career totals (Type I score ≥ 6.0) **and** outstanding career averages (Type II score ≥ 2.0). So far, between 1925 and 1990, 27 players have retired with career statistics qualifying them as Type A offensive players. Of these, every one who has appeared on a Hall of Fame Ballot has been selected for the Hall of Fame (see Table 1)--Reggie Jackson, Jim Rice, and Mike Schmidt are Type A players who have retired too recently for selection. I conclude (unremarkably) that all Type A players will make the Hall of Fame. Among active (in 1991) players, there are four whose current career statistics make them Type A players--George Brett, Dwight Evans, Rickey Henderson, and Eddie Murray.

Type B players have outstanding career *totals*, without outstanding career *averages* (Type I score ≥ 6.0; Type II score < 2.0). There are 29 retired Type B players, of whom 14 have been selected for the Hall of Fame (see Table 2) and five active Type B players. **Type C** players have outstanding career *averages*, without outstanding career *totals* (Type I score < 6.0; Type II score ≥ 2.0). There are 21 retired Type C players, of whom nine have been selected for the Hall of Fame (see Table 3) and three active Type C players.

So all Type A players make the Hall of Fame, about half the Type B players make the Hall of Fame, and about 40% of Type C players make the Hall of Fame. This leaves **Type D** players--their career totals are not outstanding (Type I score < 6.0) and their career averages are also not outstanding (Type II score < 2.0). To date, 37 out of about 250 retired Type D players have made the Hall of Fame. Among players retiring in 1925 and later, 84 non-pitchers have been selected for the Hall of Fame and 44% of them have been Type D players. Table 4 lists the Type D players who have made the Hall of Fame. There are 38 active Type D players (although some of them may change categories before they retire).

Modeling HOF Membership. Clearly, we need not concern ourselves much with Type A players. Our history says they are HOF players. We do need to examine the other types, to see whether selection standards are consistent. For this purpose, I combined Type B and Type C players together, to obtain a large enough sample to work with, and analyzed them separately from Type D players.

Modeling HOF membership is made difficult because membership is a yes-or-no phenomenon--you can't be 43% in the Hall of Fame. We cannot, therefore, use one of the standard tools that we might otherwise use to model selection (regression models). We have to account for the fact that the dependent variable--HOF membership--is what is known as a dichotomous variable (it takes on only two values--one for HOF members and zero for non-members). We should also incorporate information on positions played and on defensive performance in our analysis. I used data on games played at each position (from ***The Baseball Encyclopedia***) and calculated the percentage of games played at each position as my measure of positions played. I also used Pete Palmer's measure called "Fielder Runs" (FR) (from ***Total Baseball***) as a measure of defense. Finally, I used the Type I and Type II scores as my measures of offense.

This allowed me to estimate a model of HOF membership for Type B&C players (combined) and for Type D players. The model for Type B&C players is shown in Table 7 and the models for Type D players are shown in Table 8. In estimating these models, I used only players who had retired no later than 1975. Using players with more recent retirement dates in the estimation process runs the risk of including players with very good chances of being elected to the Hall of Fame.

For Type B&C players, defense has, apparently, not historically been a Hall of Fame qualification. Higher Type I scores and higher Type II scores increase a player's probability of selection. Players who played more at first base or in the outfield were *less* likely to be selected, given their offense, than were catchers or middle infielders. The model identifies only *one* non-HOF Type B or Type C player who retired prior to 1976 as Hall-of-Fame qualified--Ken Boyer, with an estimated 88.5% probability of making the Hall of Fame (see Table 2). Three of the Type B&C players in the Hall of Fame have predicted probabilities less than 50% of selection--John Mize (43.3%), Sam Rice (26.2%), and Hack Wilson (18.5%). Hack Wilson's selection was controversial at the time it was made.

For the Type D players, I estimated both an overall model and several position-specific models. For outfielders, first baseman, and third basemen, the overall model worked best; for catchers and middle infielders, the position-specific models worked best. I might add that the "best" models for Type D players aren't as good as the model for Type B&C players--the decisions about Type D players are more difficult, and "intangibles" may play a greater role. Specifically, the models are awful for first baseman, third basemen, and outfielders, and pretty good for catchers and for middle infielders. For the players used to estimate the model, it was correct on 28/32 Type D catchers (6/8 of those in the HOF and 22/24 of those not in the HOF), 21/25 first basemen (1/4 and 20/21), 12/16 second basemen (3/4 and 9/12), 25/27 shortstops (5/7 and 20/20), 17/21 third basemen (0/3 and 17/18), and 43/49 outfielders (3/9 and 40/40).

For catchers, the model "misses: Ernie Lombardi and Ray Schalk among HOF members and predicts that Smokey Burgess and Jim Hegan (his defense was incredible), among non-HOF catchers, have selection probabilities greater than 50%. For first basemen, the model misses Jim Bottomley, George Kelly, and Bill Terry among HOF members (getting only George Sisler right), and predicts HOF membership for Gil Hodges.

For second basemen, the model misses Tony Lazzeri among the four HOF members, and predicts HOF membership for Nellie Fox, Jim Gilliam, and Bill Mazeroski. For shortstops, the model misses Travis Jackson and Joe Sewell and predicts that none of the non-HOF shortstops in the sample will be selected.

For third basemen, the model misses all three HOF members (George Kell, Freddie Lindstrom, and Pie Traynor), predicting membership for Ron Santo instead. For outfielders, the model gets three of nine HOF members right (Max Carey, Harry Hooper, and Zack Wheat). None of the non-HOF outfielders in the sample look like plausible members.

The Type D models say that Freddie Lindstrom is the "worst" HOF selection, with only a 2% chance of being selected. Gabby Hartnett is the "best" selection, with a 99.9% chance of selection.

Overall, the models for Type D players correctly identify 18 out of 35 Hall of Fame members and 128 of 135 non-members; overall, the model correctly identifies 146 out of 170 Type D players (86%), which is fairly good for this type of model. Clearly, modeling Hall of Fame selection for Type D first basemen, third basemen, and outfielders is more difficult (4/16 HOF members correctly identified) than for catchers and middle infielders (14/19 correctly identified).

The models suggest that Smoky Burgess and Jim Hegan seem to have adequate qualifications for catchers (I disagree with the model in both of these cases); that Gil Hodges, Ron Santo, and Joe Torre (who also qualifies if we treat him as a catcher) seem to have adequate qualifications for cornermen; and that Nellie Fox, Jim Gilliam, and Bill Mazeroski have adequate qualifications for middle infielders. Fox, of course, is almost certain to be selected as soon as he can be considered by the Veterans' Committee.

Testing the Model. One way to test a model like this is to see how well it does for players who were not included in the sample used to estimate the model. In this case, we can look at two groups of players, those retiring between 1976 and 1990, and those players still active, to see which players the models predict will make the Hall of Fame. Since all the Type A players make the Hall of Fame, we can ignore them and concentrate on the Type B, C, and D players.

So far, no Type B or Type C players from the 1976-1990 period have been selected to the Hall of Fame. The model for Type B&C players suggests that 10 players from this era have qualifications similar to players previously selected. These 10 players are identified in bold-faced type in Tables 2 and 3, and there may be some surprising players in this group. I don't know how many people would have identified Jose Cruz, for example, as a potential Hall of Fame member, given the effects of the Astrodome on his statistics. Nonetheless, the model gives him a 60% change of selection. Darrell Evans, similarly, may not have seemed a particularly outstanding HOF candidate either (with a career 0.248 batting average), but his home runs, walks, and RBIs combine to boost his qualifications.

It remains to be seen how the selection process will treat players with a substantial number of games as a designated hitter (Singleton, Staub), because this is a new offensive category. It is likely that the offensive entrance requirements will be higher for DHs, however, so Singleton and Staub are definite longshots.

Two Type D players (John Bench and Brooks Robinson) who retired after 1975 have been selected to the Hall of Fame, and the models for Type D players correctly identify both of them. The models suggest that three additional catchers (Darrell Porter, Ted Simmons, and Jim Sundberg) and four additional cornermen (Buddy Bell, Bill Buckner, Toby Harrah, and Graig Nettles) have sufficient qualifications for membership. In addition, the models pick Dave Concepcion, Bobby Grich, and Jim Wynn as potential HOF members.

The models, then, suggest that there are 21 good HOF candidates, combining Type B, C, and D players, who have retired since 1976. Of these, only two--Bench and Brooks Robinson--have been selected. Of course, the other 19 are still eligible for selection, and some of them are likely to go into the Hall of Fame.

We can also look at active players (see Table 6) who have played 1500 or more games (1200 for catchers). In 1991, 50 players met these qualifications. Four were Type A players--George Brett, Dwight Evans, Rickey Henderson, and Eddie Murray. Evans may become the first Type A player *not* to make the Hall of Fame. There were five Type B players and three Type C players. The model for Type B&C players suggests that all of them except Pedro Guerrero are likely to make the Hall of Fame. Here, Jack Clark seems the real longshot (despite the model's prediction that his HOF chances are better than Andre Dawson's).

Among the 38 Type D players, the models predict HOF membership for six--Gary Carter, Willie Randolph, Ozzie Smith, Lou Whitaker, Brian Downing (and I don't believe he will make it) and Dale Murphy. At least two Type D players are likely to wind up as either Type A, B, or C players-- Cal Ripken and Tim Raines--so it may be too early to consider their chances.

Conclusions. One thing about statistical models is that one need not feel compelled to agree with their predictions. When I presented an earlier version of these models at the Chicago Emil Rothe Chapter's winter regional meeting, I commented that the HOF selectors apparently thought the fact that Earle Combs played *with* Babe Ruth meant he played *like* Babe Ruth. (I tend to agree with the model about Combs.) Other attendees at the meeting pointed to his 0.325 lifetime BA (about 1.25 standard deviations above average--and he had average power and average ability to draw a walk) and to his defense [a fairly large number of PO per game (but virtually no assists--69 total, according to the ***Baseball Encyclopedia***, for his *career*)] to suggest that he was a valid HOF member.

I personally disagree with the model in several cases--I don't think Smoky Burgess or Jim Hegan or Brian Downing or Toby Harrah or Rusty Staub or Ken Singleton belongs in the HOF; I'm not sure about Gil Hodges or Willie Randolph or Bobby Bonds or Darrell Evans or Fred Lynn. But their numbers are consistent with past selections, so the model says they are qualified.

Finally, of course, we must contend with the changes in the rules, particularly the rules governing consideration by the Veterans' Committee. In the past, the Veteran's Committee could correct such blatantly ignorant votes as that on Bobby Grich (with the potential cost, of course, that the Veterans'

Committee might someday put Ken Keltner in the HOF). Now, Grich--who is as qualified as Nellie Fox or Bill Mazeroski--cannot ever be considered again for the Hall of Fame, unless the rules change again. Such, I suppose, is life.

Table 1: Truly Great Offensive Players (Type A Players)
Hank Aaron
Ty Cobb
Eddie Collins
Joe Dimaggio
Jimmie Foxx
Lou Gehrig
Harry Heilman
Rogers Hornsby
Al Kaline
Mickey Mantle
Eddie Mathews
Willie Mays
Willie McCovey
Joe Morgan
Stan Musial
Mel Ott
Reggie Jackson*
Jim Rice*
Frank Robinson
Babe Ruth
Mike Schmidt*
Duke Snider
Tris Speaker
Willie Stargell
Billy Williams
Ted Williams
Carl Yastrzemski

*These players are not yet in the Baseball Hall of Fame.

Table 2: Great Offensive Players With Outstanding Career Totals (Type B Players)

Player	Probability of Making HOF
Luke Appling*	93.9%
Richie Ashburn	44.4%
Ernie Banks*	50.2%
Bobby Bonds	**68.9%**
Ken Boyer	**88.5%**
Lou Brock*	73.0%
Rod Carew*	85.4%
Cesar Cedeno	47.0%
R. Clemente*	66.0%
Jose Cruz	**60.1%**
Willie Davis	13.7%
Darrell Evans	**85.8%**
F. Frisch*	81.5%
C. Gehringer*	93.8%
Goose Goslin*	84.0%
Joe Kuhel	8.8%
Joe Mediwck*	55.3%
Wally Moses	34.0%
Al Oliver	36.6%
Tony Perez	**66.9%**
Vada Pinson	37.8%
PeeWee Reese*	70.9%
*Sam Rice**	*26.2%*
Pete Rose	**99.1%**
Al Simmons*	58.9%
E. Slaughter*	74.6%
Rusty Staub	**71.3%**
Mickey Vernon	15.0%
Paul Waner*	66.5%

*In the Hall of Fame.
BOLD: Not in HOF; the model predicts a probability > 0.5 of membership.
ITALIC: In the HOF; the model predicts a probability < 0.5 of membership.

Table 3: Great Offensive Players With Outstanding Career Averages (Type C Players)	
Player	Probability of Making HOF
Dick Allen	**76.8%**
Earl Averill*	54.7%
Rico Carty	25.6%
Norm Cash	10.6%
M. Cochrane*	62.4%
Lary Doby	29.5%
Jake Fournier	6.5%
H. Greenberg*	83.0%
K. Hernandez	23.1%
H. Killebrew*	82.0%
Ralph Kiner*	69.5%
Greg Luzinski	49.4%
Fred Lynn	**65.0%**
*John Mize**	*43.3%*
J. Robinson*	59.1%
Ken Singleton	**59.8%**
Reggie Smith	**64.5%**
Gene Tenace	13.5%
A. Thornton	17.9%
Arky Vaughan*	92.0%
*Hack Wilson**	*18.5%*

*In the Hall of Fame.
BOLD: Not in HOF; the model predicts a probability > 0.5 of membership.
ITALIC: In the HOF; the model predicts a probability < 0.5 of membership.

Table 4: Hall-of-Fame Members and HOF Membership Probability (Type D Players)	
Player	Probability of Making HOF
Catchers	
Johnny Bench	99.9%
Yogi Berra	92.1%
Roy Campanella	83.6%
Bill Dickey	80.3%
Rick Ferrell	81.3%
Gabby Hartnett	99.9%
Ernie Lombardi	*25.9%*
Al Lopez	87.7%
Ray Schalk	*27.8%*
Corners	
Jim Bottomley	*22.7%*
George Kell	*8.8%*
George Kelly	*5.0%*
Freddie Lindstrom	*2.0%*
Brooks Robinson	90.3%
George Sisler	56.1%
Bill Terry	*26.4%*
Pie Traynor	*23.8%*
Middle Infield	
Luis Aparicio	83.3%
Davy Bancroft	52.1%
Lou Boudreau	55.7%
Joe Cronin	75.3%
Bobby Doerr	99.9%
Billy Herman	95.1%
Travis Jackson	*12.5%*
Tony Lazzeri	*28.8%*
Rabbit Maranville	81.1%
Red Schoendienst	99.8%
Joe Sewell	*42.3%*
Outfield	
Max Carey	87.6%
Earle Combs	*6.9%*
Kiki Cuyler	*47.9%*
Harry Hooper	68.9%
Chuck Klein	*21.7%*
Heinie Manush	*26.0%*
Edd Roush	*18.8%*
Lloyd Waner	*5.0%*
Zack Wheat	68.7%

ITALIC: In the HOF; the model predicts a probability < 0.5 of membership.

Table 5: Type D Players Not in HOF, with a Probability > 0.5 of Membership and Other Players of Interest

Player	Probability of Making HOF
Catcher	
Bob Boone	41.3%
Smoky Burgess	**83.9%**
Bill Freehan	7.9%
Jim Hegan	**52.4%**
Elston Howard	1.2%
Thurman Munson	2.6%
Darrell Porter	**57.0%**
Ted Simmons	**89.5%**
Jim Sundberg	**95.0%**
Steve Yeager	0.8%
Corners	
Buddy Bell	**84.7%**
Bill Buckner	**70.5%**
Ron Cey	45.3%
Bob Elliott	38.4%
Steve Garvey	35.1%
Toby Harrah	**83.6%**
Gil Hodges	**60.0%**
Ken Keltner	6.0%
Graig Nettles	**87.1%**
Ron Santo	**83.5%**
Joe Torre	**59.5%**
Middle Infield	
Dave Concepcion	**54.0%**
Leo Durocher	0.7%
Nellie Fox	**99.9%**
Jim Gilliam	**99.9%**
Joe Gordon	33.9%
Bobby Grich	**96.3%**
Bill Mazeroski	**84.9%**
Vern Stephens	25.1%
Frank White	22.0%
Maury Wills	13.7%

Table 5: Type D Players Not in HOF, with a Probability > 0.5 of Membership and Other Players of Interest

Player	Probability of Making HOF
Outfield	
Don Baylor	48.2%
Rocky Colavito	40.2%
George Foster	44.4%
Bob L. Johnson	42.9%
Chet Lemon	39.8%
Garry Matthews	37.1%
Hal McRae	42.1%
Minnie Minoso	49.4%
Amos Otis	48.9%
Jim Wynn	**51.2%**

Table 6: Active Players (1991) and Probabilities of Making HOF

Player	Probability of Making HOF
Type A Players	
George Brett	
Dwight Evans	
Rickey Henderson	
Eddie Murray	
Type B Players	
Andre Dawson	58.4%
Carlton Fisk	84.6%
Dave Parker	77.4%
Dave Winfield	92.4%
Robin Yount	93.2%
Type C Players	
Wade Boggs	94.5%
Jack Clark	71.0%
Pedro Guerrero	18.1%

Table 6 (Continued)	
Type D Players	
Catchers	
Gary Carter	**96.8%**
Rick Cerone	0.0%
Rick Dempsey	44.9%
Ron Hassey	2.4%
Mike Heath	0.0%
Terry Kennedy	0.0%
Lance Parrish	8.7%
Tony Pena	0.5%
Mike Scioscia	19.1%
Ernie Whitt	14.1%
Middle Infield	
Jim Gantner	0.0%
Alfredo Griffin	2.9%
Tom Herr	0.0%
Rafael Ramirez	0.8%
Willie Randolph	**99.3%**
Cal Ripken	35.9%
Steve Sax	0.2%
Ozzie Smith	**75.9%**
Gary Templeton	24.1%
Alan Trammell	34.5%
Lou Whitaker	**95.7%**
Corners	
Hubie Brooks	0.8%
Gary Gaetti	4.3%
C. Lansford	1.9%
Paul Molitor	31.2%
Ken Oberkfell	1.6%
Tim Wallach	6.9%
Outfield	
Harold Baines	12.8%
Tom Brunansky	6.0%
Brett Butler	9.4%
Chili Davis	5.3%
Brian Downing	**62.7%**
Ken.Griffey Sr.	36.7%
Lloyd Moseby	5.5%
Dale Murphy	**66.3%**
Terry Puhl	2.7%
Tim Raines	34.8%
Willie Wilson	17.2%

BOLD: Model predicts a probability > 0.5 of making the HOF.

Table 7: Modeling HOF Membership for Type B and Type C Players

Variable	Coefficient
Type I Score	+0.288 (2.43)
Type II Score	+0.711 (2.21)
Percent of Games at First Base	-4.002 (-2.20)
Outfield	-1.843 (-1.56)
Constant	-0.480 (-0.42)
Percent Correct	
Total	80.9%
HOF = 1	87.5%
HOF = 0	72.2%

Table 8: Models of Hall of Fame Membership for Type D Players

Variable	Total	Catchers	Middle Infield
Type I	+0.228 (1.77)	-0.963 (-1.81)	
Type II	+0.345 (1.44)	+2.943 (2.34)	+1.284 (2.62)
FR	+0.008 (2.43)	+0.047 (2.29)	+0.009 (1.88)
Games	+0.002 (1.34)	+0.015 (2.34)	+0.006 (2.49)
Percent of Games at Catcher	+3.193 (3.40)		
Shortstop	+1.129 (1.44)		
Constant	-5.149 (-2.03)	-29.019 (-2.30)	-11.139 (-2.60)